AFRL-OSR-VA-TR-2013-0564

(CONGRESSIONAL) HIGH ASSURANCE SOFTWARE

WILLIAM MAHONEY

UNIVERSITY OF NEBRASKA

10/22/2013
Final Report

DISTRIBUTION A: Distribution approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)/RSL
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* 16-10-2013	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* 6/15/2007-9/30/2013

4. TITLE AND SUBTITLE

(Congressional Interest) High Assurance Software

5a. CONTRACT NUMBER

N/A

5b. GRANT NUMBER

FA9550-07-1-0499

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Mahoney, William R.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Board of Regents, University of Nebraska d/b/a University of Nebraska at Omaha
Eppley Administration Building, Room 203
6001 Dodge Street, Omaha, NE 68182-0116

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

USAF, AFRL, AF Office of Scientific Research
875 North Randolph Street, Suite 325, Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR; AA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

N/A

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A - Approved for Public Release

13. SUPPLEMENTARY NOTES

N/A

14. ABSTRACT

The High Assurance Software funding source was used to conduct advanced research in several areas related to software and security. The grant was used primarily in five research areas: the study of software vulnerabilities by the use of semantic templates, the study of software reliability relative to the release cycle and architecture of the source, prediction of cyber attacks based on the analysis of open-source intelligence, examination of computer viruses using bioinformatics tools, and the anticipation and avoidance of intellectual property theft through software obfuscation.

15. SUBJECT TERMS

semantic templates, software reliability, obfuscation, computer viruses

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU		**19b. TELEPHONE NUMBER** *(include area code)*

Progress Report for the Nebraska University Consortium on Information Assurance
For the period of June 15, 2007 to September 3, 2013
For US Department of Defense /Air Force Office of Scientific Research
In Regards To FA9550-07-1-0499 High Assurance Software

Report Type
Final Performance Report

Primary Contact E-mail
wmahoney@unomaha.edu

Primary Contact Phone Number
(402) 554-3975

Organization / Institution name
University of Nebraska at Omaha

Award Information

Grant/Contract Title
High Assurance Software

Grant/Contract Number
DOD

Principal Investigator Name
Dr. William Mahoney

Program Manager
Dr. Robert Herklotz

Report Information - Annual Report

Reporting Period Start Date
June 15, 2007

Reporting Period End Date
September 30, 2013

Report Abstract:

Abstract The High Assurance Software funding source was used to conduct advanced research in several areas related to software and security. In particular, we had several ongoing research efforts in the following areas: Architecture-based Software Reliability, the use of Semantic Templates to study Software Vulnerabilities, a project called CyCast which is an attempt to predict cyber attacks on software systems, CyberGenome, which is utilizing local expertise in bioinformatics to examine not nucleotides, but software components, and automatic software obfuscation.

Over the course of the past reporting year, we have published 24 reports on these projects and have two more in the pipeline. While none of these have been in the CyberGenome area, we continue to think that the advanced techniques brought to bear from the biological analog will be of use in such areas as file attribution. In our previous annual report we described a publication – Pocket guide for Software Assurance Workforce, Training and Education, Eds. Gandhi, R., Department of Homeland Security (DHS) Software Assurance Initiative, 2010. This publication continues to grow in status within the DHS and other communities and the High Assurance Software funding has also been used for travel to DHS events.

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Additional Information

Archival Publications (published) during reporting period: Details of Research

Semantic Template Guided Rule Authoring for Weakness Detection Masters Thesis. Suhrit Rimal.

Helping Programmers Understand and Study Software Security Weaknesses: Semantic Templates Gandhi, R. A., Presentation at the SwA Working Group Sessions - MITRE-1, McLean, VA,

Bridging to the Future – Emerging Trends in Cybersecurity Gandhi, R. A., Presentation at the 24th FISSEA's 24th Annual Conference: March 15 - 17, 2011, National Institute of Standards and Technology Gaithersburg, Maryland.

Using Semantic Templates to Study Vulnerabilities Recorded in Large Software Repositories W. Yan, R.A. Gandhi, and H. Siy, (2010) In Proc. of The 6th International Workshop on Software Engineering for Secure Systems (SESS'10) at the 32nd International Conference on Software Engineering (ICSE 2010), Cape Town, South Africa.

Studying Software Vulnerabilities R.A. Gandhi, H. Siy, and W. Yan, (2010) CrossTalk, The Journal of Defense Software Engineering.

Optimal Values for Disrupting x86-64 Reverse Assemblers Sara Shinn, William Mahoney, International Journal of Computer Science and Network Security, Volume 11, Number 11, November 2011.

Semi-Automatic Annotation of Natural Language Vulnerability Reports. Wu, Y., Gandhi, R., & Siy, H. (2013). *International Journal of Secure Software Engineering (IJSSE), 4*(3), 18-41. doi:10.4018/jsse.2013070102

A Social Dimensional Cyber Threat Model with Formal Concept Analysis and Fact-Proposition Inference, Sharma, A., Gandhi, R.A., Mahoney, W., Sousan, W., Zhu, Q., (2013) International Journal of Information and Computer Security.

Lightweight formalization of security weaknesses, Gandhi, R.A., Siy, H., Yan, Wu, (2013) Workshop on Formal Methods in Software Engineering (FormaliSE): 25 May 2013, San Francisco (USA), in conjunction with International Conference on Software Engineering. 2013.

Fingerprinting Malware Using Bioinformatics Tools Building a Classifier for the Zeus Virus, Pedersen, J., Bastola D., Dick, K., Gandhi, R., Mahoney, M., (2013) International Conference on Security and Management (SAM'13), Las Vegas, USA, July 22 - 25.

Early Security Patterns: A Collection of Constraints to Describe Regulatory Security Requirements, Gandhi, R.A., Rahmani, M. RePa 2012 : Second International Workshop on Requirements Patterns, IEEE International Conference on Requirements Engineering, Chicago, Sept, 2012

BLAST Your Way through Malware: Malware Analysis Assisted by Bioinformatics Tools, Pedersen, J., Bastola D., Dick, K., Gandhi, R., Mahoney, M., The 2012 International Conference on Security and Management (SAM'12), Las Vegas, USA, July 16 - 19.

Architecture-based Reliability Modeling of Web Services in Multilayer Environment, C. Rahmani, A. Azadmanesh, H. Siy, Principles of Engineering Service-Oriented Systems (PESOS), Int'l Conference on Software Engineering (ICSE) , Hawaii, May 2011.

Architecture-based Reliability Modeling of Web Services Using Petri Nets, A. Azadmanesh, H. Siy, 12th IEEE International High Assurance Systems Engineering Symposium (HASE), San Jose, USA, Nov 2010.

A Comparative Analysis of Open Source Software Reliability, C. Rahmani, A. Azadmanesh. Najjar, Journal of Software, 5(12), 2010.

Exploratory Failure Analysis on Open Source Software, IEEE Int'l Conference on Software Technology and Engineering (ICSTE), C. Rahmani, S. Srinivasan, A. Azadmanesh, San Juan, Puerto Rico, Oct 2010.

A Study on Defect Density of Open Source Software, IEEE /ACIS Int'l Conference on Computer and Information Science, C. Rahmani, D. Khazanchi, Yamagata, Japan, Aug 2010.

The Cultural, Social, Economic, and Political Dimensions of Cyber Attacks, IEEE Technology and Society Magazine, Vol. 30, No. 1, Spring 2011 Issue. Gandhi, R.A., Sharma, A. Mahoney, W., Susan, W., Quiming, Z., Laplante, P.

Building a Social Dimensional Threat Model from Current and Historic Events of Cyber Attacks, International Symposium on Secure Computing (SecureCom-10) In conjunction with The Second IEEE International Conference on Privacy, Security, Risk and Trust, Sharma, A., Gandhi, R.A., Mahoney, W., Sousan, W., Zhu, Q.

Exploring Social Contexts along the Time Dimension: Temporal Analysis of Named Entities, International Symposium on Secure Computing (SecureCom-10) In conjunction with The Second IEEE International Conference on Privacy, Security, Risk and Trust, Walnez, B., Gandhi, R.A., Mahoney, Zhu, Q.,

Using Term Extraction Patterns to Discover Coherent Relationships from Open Source Intelligence, International Symposium on Secure Computing (SecureCom-10) In conjunction with The Second IEEE International Conference on Privacy, Security, Risk and Trust, W., Sousan, Gandhi, R.A., Mahoney, W., Zhu, Q., Sharma, A.,

Studying Security Vulnerabilities, CrossTalk, The Journal of Defense Software Engineering, Sept/Oct issue 2010. Gandhi, R.A., Siy, H., Wu, Y.

Empirical Results on the Study of Software Vulnerabilities (NIER Track). In proceedings of the 33rd International Conference on Software Engineering (ICSE 2011), Waikiki, Honolulu, Hawaii, May 21-28, 2011, Siy, H., Wu, Y., Gandhi, R.A.

Publications submitted:

Modifications to GCC for Anti-Reverse Engineering W. Mahoney, (2012), IET Information Security. Under review.

Semantic Relevance Analysis of Subject-Predicate-Object (SPO) Triples, Ranjana Kumar, Robin Gandhi, William Mahoney, Parvathi Chundi, and Quiming Zhu, submitted to International Journal on Software and Knowledge Engineering.

Visual Analytics for Software Weaknesses, Tahmasbi, N., Gandhi, R., Siy, H., Submitted for conference publication

Changes in research objectives (if any): None

Change in AFOSR Program Manager, if any: None

Extensions granted or milestones slipped, if any: None

Include any new discoveries, inventions, or patent disclosures during this reporting period (if none, report none): None

We continue to utilize the funding for four concurrent research projects, which are detailed below. Prior annual reports also listed start-up research in SCADA security; this project has been moved to a different funding source.

- **Architecture-Based Software Reliability**: A white-box approach which attempts to measure the quality of a software system based on its structure. Nationally, this may allow companies within the United States to know, based on the implementation methods for a complex system, what trust level should be associated with that system. The research in software reliability has expanded into web services. Web services, such as electronic shopping and banking services, have permeated our daily lives due to the ease of use, flexibility and reduction of cost in providing the services. In this regard, the research approaches are focused on 1) transforming the architecture of the web-service systems to gray layers, where a gray layer represents collection of components representing some logical behavior of the system, and 2) the behavior of web-service systems are simulated using Petri net models. The models are validated by conducting empirical studies and injecting the data collected into the Petri nets. The questions this research attempts to answer include the following: How can a web-service architecture be extracted? How can reliability and security of SOA be analyzed? How to predict failures? What are the effects of application server failures on customer satisfaction? How to create and validate Petri net models to mock the web-service system behavior? How to create flexible Petri net models that can easily be changed based on future needs.

Accomplishments: Graduate (PhD) student Cobra Rahmani, in conjunction with Dr. Azad Azadmanesh and others, has published five research articles within the last year. Currently, there is another article in progress. These achievements have been recognized by the University of Nebraska-Omaha by presenting a university wide dissertation award to Ms. Rahmani. Additionally, the maturity of research in software reliability has facilitated the creation of a Software Reliability Group (SRG). Currently, a number of students have shown interest in joining the group (http://ahvaz.ist.unomaha.edu/srg/).

- **Using Semantic Templates to study Vulnerabilities from Large Software Repositories**: To cope with growing software complexity, programmers need better mental models to sense the possibility of vulnerability. There is no shortage of weakness enumerations and categorization, but they are not in a form that facilitates human understanding and recall. For example, the Common Weakness Enumeration (CWE) contains over 50 highly inter-related weakness definitions just to comprehend the possibility of buffer overflows. In this research, we have developed semantic templates to help programmers study software vulnerabilities and avoid them during software development. Using Semantic Templates experiments indicate a definite improvement in the programmer ability to understand the CWEs related to the underlying software fault, weakness characteristics, resources and locations affected and the consequences of a given vulnerability. We have also developed capabilities to semi-automatically annotate vulnerability descriptions with CWEs using text-mining tools. We have also conducted research in the visualization of the relationships between the weakness standards. Most recently, we have developed lightweight formalizations of software weaknesses and regulatory requirements to reduce ambiguity in their natural language specification.

Accomplishments: PhD candidate Yan Wu, along with Dr. Robin Gandhi and Dr. Harvey Siy, have published and presented their work at the 2011 International Conference on Software Engineering, in the new and emerging ideas track. They also published part of this work in CrossTalk: The Journal of Defense Software Engineering. Dr. Gandhi most recently presented this work at the DHS Software Assurance Working Group held at MITRE in Mclean, Virginia. Yan Wu defended her dissertation in January 2012 and is currently working at the National Institute of Standards and Technology (NIST) as a visiting researcher. She continues to work on the topic of formalization of software weaknesses identified in CWEs. Most recently, semi-automated annotation of vulnerability reports have been published in the International Journal of Secure Software. Formalization of software weaknesses and regulatory requirements has been published and presented in highly competitive workshops at the International Conference on Software Engineering and the International Conference on Requirements Engineering.

- **The Social, Political, Economic and Cultural (SPEC) aspects of a Cyber Attack (CyCast Project)**: James Jinhua Zhang's research: Ontology Based Document Categorization and Automatic Determination of Domain Relevance.
  The research on Ontology Based Document Categorization and Automatic Determination of Domain Relevance addresses the problem of automatically detecting the documents within some specific Domain of Interest (DoI), and extending the discovery to additional information sources to finding the topic relevant documents. The system will identify documents around a set of relevant concepts expressed in an ontology, such that it is possible to classify a given document (Web page, blogs, newsprint and micro-blogs, tweets, and social network contents) into a specific category, and to find documents with specific theme, domain, or topic categorization from a set of given documents.
  Features of the approach include: 1) Applying machine learning techniques to derive a set of Term frequency inverse document frequency (tf-idf) weights from a given set of document corpus according to a given set of textual concepts in a specified ontology. 2) The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus, i.e., the tf-Idf weights is a measure of the general importance of the term in the ontology with respect to the domain of interest. 3) The tf-idf weights, once derived from the learning corpus, are then used to evaluate how important a word is to a document with respect to a specific category in a collection or corpus. 4) The system works with knowledge from a given ontology which is built by a statistical relational learning (the tf-idf learning) process applied to a given corpus of documents in a domain of interest. The learning process derives a set of tf-idf weights that are used to evaluate the contents of documents.
  The research will deal with the Document Categorization and its application to automated classification of news articles relevant to cyber attacks into social, political, economic, and cultural (SPEC) domains with the use of a highly organized and inter-linked vocabulary (ontology) of SPEC dimensional cyber attacks and a tf-idf weighting technique. The resulting system can be used to obtain the SPEC factors by extraction of Web resources such as news articles, and to apply a formal model of analytic inference in a decision-tree structure to compute the likelihood of impending cyber attacks in SPEC dimensions. Applying to the cyber security domain for compute the likelihood of impending cyber attacks in SPEC dimensions, the document categorization system will allow the quest to make sense of processed documents and unfold cyber security relevant events in SPEC hotspots of the world. Applying the automatic categorization, the resulting cyber attack forecasting system (CyCast) will be able to connect the dots from the SPEC-related event reports and news stories to provide users an integrated picture of situation awareness for the cyber security domain of their concern.

Accomplishments: Several technical advancements have been disseminated in three papers at the workshops at the IEEE International Conference on Privacy, Security, Risk and Trust. Our web site is live and allows our users to search cyber attacks using the SPEC dimensions of our research (http://kewi.unomaha.edu). Several web applications are in the process of going live for more interactive experience. One journal paper will be published in 2013 and one journal paper is under review.

- **CyberGemone**: The primary goal of this project is to determine information about the origin of a computer artifact by analyzing its content using Bioinformatics tools. A secondary goal is to analyze the evolution of a piece of software over time, i.e., to try to characterize the amount of change that a piece of software has had between versions.

  The first research task was converting any given computer artifact into a format usable by the Bioinformatics tools. This was done, and any computer artifact (any file) can be converted into a FASTA format file, which can then be processed by Bioinformatics tools. The content of the file is a nucleotide representation of the original artifact file. This process can be reversed to obtain the original file from the FASTA file. We then use Bioinformatics tools to look for elements common to all files of a particular type, using Restriction Enzymes to try to categorize computer artifacts by the number of "DNA cuts". A heat map-clustering algorithm is used along with the results of the cut rates for various enzymes and data files. The eventual goal of the research is to determine whether we can correctly ascertain the attribution of a file in this way.

The second objective of the research is to determine whether these tools are useful for tracking the evolution of computer viruses. At this time the research for this second goal has not been started.

Accomplishments: Our findings from this project have been disseminated in two papers at the International Conference on Security and Management. We are also starting to receive correspondence from other research groups that are interested in using our techniques for malware analysis.

- **Automatic Obfuscation**: Since we actively teach a reverse engineering course in our IA curriculum it is natural that a research area is how to thwart the reverse engineering process. Towards this end we have conducted work, which has yielded a version of the Gnu Compiler Collection that automatically obfuscates the object code created in the compilation process.

Accomplishments: We have one published paper on this work and are awaiting an additional papers acceptance.   We have sent whitepapers to various funding sources to push this research forward using more modern tool chains and processors.

Partially sponsored by this research funding:

Dr. Robin Gandhi – Associate Professor, Information Assurance
Ms. Connie Jones – Outreach Coordinator / Program Development, College of IS&T
Mr. Charles Spence – Student Lab Manager
Ms. Cobra Rahmani – PhD Graduate Student, Architecture-Based Software Reliability project.
Ms. Ruchi Sodhani – Graduate Student, Enterprise Risk Management project.
Ms. Yan Wu – PhD Graduate Student, Large Software Repositories project.
Mr. James Jinhua Zhang – Graduate Student, CyCast Project

Associated with the research efforts:

Dr. Azad Azadmanesh – Professor, Computer Science, Architecture-Based Software Reliability project
Dr. Dhundy (Kiran) Bastola – Associate Professor, Bioinformatics, CyberGenome project
Dr. Ken Dick – Senior Research Fellow, CyberGenome project
Dr. Bill Mahoney – Associate Professor, Information Assurance, All projects
Dr. Harvey Siy – Associate Professor, Computer Science, Large Software Repositories project
Mr./Dr. William Sousan – PhD Graduate, CyCast project
Dr. Quiming Zhu – Professor, Computer Science, CyCast Project

End of report

Best Regards,

William Mahoney, PhD
Executive Director,
Nebraska University Consortium for Information Assurance, (NUCIA)
College of IS&T
Peter Kiewit Institute, 282F
University of Nebraska, Omaha
1110 S. 67th Street
Omaha, NE 68182-0694
1-402-554-3975 office
Alternative email: Is there another email besides wmahoney@unomaha.edu ?
This report also sent to: